

A Variance-Minimizing Filter for Large-Scale Applications

P. J. VAN LEEUWEN

IMAU, Utrecht University, Utrecht, Netherlands

(Manuscript received 25 June 2002, in final form 2 January 2003)

ABSTRACT

A truly variance-minimizing filter is introduced and its performance is demonstrated with the Korteweg–DeVries (KdV) equation and with a multilayer quasigeostrophic model of the ocean area around South Africa. It is recalled that Kalman-like filters are not variance minimizing for nonlinear model dynamics and that four-dimensional variational data assimilation (4DVAR)-like methods relying on perfect model dynamics have difficulty with providing error estimates. The new method does not have these drawbacks. In fact, it combines advantages from both methods in that it does provide error estimates while automatically having balanced states after analysis, without extra computations. It is based on ensemble or Monte Carlo integrations to simulate the probability density of the model evolution. When observations are available, the so-called importance resampling algorithm is applied. From Bayes's theorem it follows that each ensemble member receives a new weight dependent on its "distance" to the observations. Because the weights are strongly varying, a resampling of the ensemble is necessary. This resampling is done such that members with high weights are duplicated according to their weights, while low-weight members are largely ignored. In passing, it is noted that data assimilation is not an inverse problem by nature, although it can be formulated that way. Also, it is shown that the posterior variance can be larger than the prior if the usual Gaussian framework is set aside. However, in the examples presented here, the entropy of the probability densities is decreasing. The application to the ocean area around South Africa, governed by strongly nonlinear dynamics, shows that the method is working satisfactorily. The strong and weak points of the method are discussed and possible improvements are proposed.

1. Introduction

In areas of the World Ocean, where even state-of-the-art numerical models still have serious shortcomings, data assimilation can be of use. Smoothers, in which observations of the system are not only used for future evolution but also back in time, can be especially useful for pointing to missing model physics, problematic forcing, or wrong boundary conditions.

Another area where data assimilation is important is forecasting. By providing an initial condition as close as possible to observations, while still allowing information from model dynamics, more accurate predictions can be made. In this case, it does not make sense to use a smoother because it contains no extra information compared to a filter, even for nonlinear models (see Evensen and van Leeuwen 2000).

The problems we are facing today regarding data assimilation are, in large part, attributable to poorly known model behavior. We all agree that all observations of a system should be accompanied by an estimate of how accurate the measurement is. For numerical models, this is not common practice. Still, numerous studies appear

in literature in which one single (climate) model run is analyzed in some detail followed by exclamations on the behavior of the model in general, or even on the real world! Obviously, one needs to know the model sensitivity to various parameters and parameterizations before these kinds of statements can be made.

While for linear problems a Kalman filter will provide the variance-minimizing solution, for nonlinear problems this is not the case (see, e.g., Jazwinski 1970). The so-called extended Kalman filter designed for nonlinear problems can only handle weakly nonlinear model behavior because the assumption is made that the error evolution evolves according to the tangent linear model evolution. Furthermore, it is assumed that the central forecast, and so the model evolution, assuming no errors in the dynamical equations, is the optimal state. The advantage of this assumption is that the optimal state evolves independently from the rest of the probability density of the model, so that the covariances are only needed at analysis times. This has led to several methods to approximate the error covariances at analysis times, avoiding the time-consuming propagation of the covariance fields. Obviously, when the model is strongly nonlinear, problems regarding the optimality of the solution arise.

The problems are, in fact, threefold. First, the evolution of the error covariances is not according to the

Corresponding author address: P. J. van Leeuwen, IMAU, Utrecht University, P.O. Box 80005, 3508 TA Utrecht, Netherlands.
E-mail: leeuwen@phys.uu.nl

tangent linear model evolution. Several ensemble methods based on either a root-mean-square approximation of the error covariance (e.g., Verlaan and Heemink 1997) or a Monte Carlo approach solve this problem. The second problem is that the central forecast is not optimal between analyses for a nonlinear model evolution. Methods like the ensemble Kalman filter (EnKF; Evensen 1994; see also Burgers et al. 1998) address this problem by propagating an ensemble of model states forward in time and taking (correctly) the variance-minimizing state as the mean of the ensemble. Interesting variants of the EnKF are presented by Heemink et al. (2001), who show that a combination of a square root filter and the EnKF is superior to either variant separately, and by Anderson (2001), who developed an ensemble adjustment Kalman filter in which no perturbation of observations is needed. Finally, the Kalman update itself is not variance minimizing (see, e.g., Jazwinski 1970). In this paper, a possible solution to this last problem is presented, using ensemble integrations to solve the first two.

The method presented here remains rather close to Bayes equation, which is at the heart of the data assimilation problem. It is shown that a frequency (or particle) representation of the probability density of the model evolution leads to a weighting of the ensemble members (or particles) related to their distance to the observations. Van Leeuwen and Evensen (1996) tried to apply this method with a smoother over a 100-day interval in a quasigeostrophic ocean model, but it failed to work. The failure is due to the fact that only very few ensemble members have a large enough weight over the 100 days, so the effective ensemble size becomes too small. However, when the weighted probability density is resampled every now and then, the idea can be made to work for quite large state spaces, as is shown in this paper. Several similar methods have been proposed for these kinds of problems (e.g., Anderson and Anderson 1999; Miller et al. 1999; Pham 2001), but they have only been tested for low-dimensional problems like the Lorenz63 (Lorenz 1963) model (although the application to large-scale models is discussed in these papers).

In the next section, the new method is outlined. Section 3 discusses an application to a simple nonlinear model that goes beyond the Lorenz equation, while still being tractable. This application on the Korteweg–DeVries (KdV) equation is compared to results obtained with the ensemble Kalman filter. Section 4 discusses a real application to the ocean area around South Africa, modeled with a five-layer quasigeostrophic model. Section 5 summarizes the paper and discusses the potential of the method for even larger-scale applications.

2. A variance-minimizing filter

In this section, a truly variance-minimizing filter and smoother are derived using Bayes's theorem. As the name suggests, a variance-minimal estimate is that es-

timate with minimal variance. Since the most commonly used way to characterize the accuracy of an estimate is to look at its variance, and because it is a sensible measure of the spread of a probability density, it makes sense to try to find this estimate. At the heart of nonlinear data assimilation lies the notion of combining probability densities of model and observations. By expressing the problem in terms of probability density functions, a Bayesian estimation problem can be formulated. In Bayesian statistics, the unknown model evolution ψ is viewed as the value of a random variable $\underline{\psi}$. The density $f_m(\psi)$ of ψ is obtained from the model somehow and is called the prior probability density. Using the definition of a conditional probability density we can derive the new, or posterior, probability density of ψ given the observations \mathbf{d} :

$$f_m(\psi|\mathbf{d}) = \frac{f_d(\mathbf{d}|\psi)f_m(\psi)}{\int f_d(\mathbf{d}|\psi)f_m(\psi) d\psi}. \quad (1)$$

The first factor in the numerator, the density $f_d(\mathbf{d}|\psi)$, is the probability density of the observations, given that the model random variable $\underline{\psi} = \psi$. The second factor is the a priori model density $f_m(\psi)$. The denominator is the probability density of the observations, written as a marginal density of the joint density of model and observations. Obviously, this is just a normalization term [see, e.g., van Leeuwen and Evensen (1996) for details].

What first comes to mind when considering this equation is that data assimilation is not an inverse problem from the start. It can be put in that form, but defining it in terms of combining probability densities is more natural. Indeed, this has direct bearing on the general idea of data assimilation: we try to combine information from observations with dynamical information in the form of partial differential equations of the underlying physics. It can be put into an inverse problem for practical purposes, but there is no need to do so in principle. As we will see, the method presented here will not be turned into an inverse problem, so no matrices have to be inverted.

As is well known, the variance-minimizing model evolution is equal to the mean of the posterior probability density:

$$\bar{\psi} = \int \psi f_m(\psi|\mathbf{d}) d\psi. \quad (2)$$

A frequency (or particle) interpretation of this equation leads to

$$\bar{\psi} = \sum_{i=1}^N \psi_i f_m(\psi_i|\mathbf{d}), \quad (3)$$

which can be rewritten as

$$\bar{\psi} = \frac{\sum_{i=1}^N \psi_i f_m(\mathbf{d} | \psi_i)}{\sum_{i=1}^N f_m(\mathbf{d} | \psi_i)}. \quad (4)$$

The meaning of this equation is that each ensemble member (or particle) is weighted by its “distance” to the observations, with the weights given by

$$w_i = \frac{f_d(\mathbf{d} | \psi_i)}{\sum_{i=1}^N f_d(\mathbf{d} | \psi_i)}. \quad (5)$$

The “distance” is found from the probability density of the observations. If the observational errors are Gaussian distributed, and, for simplicity, uncorrelated with standard deviations σ , the weights are found as

$$w_i = \frac{1}{A} \exp \left[-\frac{(\mathbf{d} - H\psi_i)^2}{2\sigma^2} \right], \quad (6)$$

in which the normalization constant A is given by

$$A = \prod \exp \left[-\frac{(\mathbf{d} - H\psi_i)^2}{2\sigma^2} \right], \quad (7)$$

and in which H is the measurement operator, which can be strongly nonlinear.

From the above, it becomes clear that all operations that have to be performed to obtain the variance-minimizing solution are direct; that is, no inversions are present. One of the advantages is that the measurement operator can be extremely difficult to linearize without damaging the calculation: we just take it as it comes. Another is that no matrices have to be inverted, either directly or iteratively. It is also noted that higher-order moments are extremely easy to obtain as soon as we have the ensemble. For instance, any moment $g(\psi)$ is obtained as

$$\overline{g(\psi)} = \sum_{i=1}^N w_i g(\psi_i), \quad (8)$$

with the weights w_i given above. Again, no inversions needed.

Now, as mentioned in the introduction, van Leeuwen and Evensen (1996) have tried to apply these ideas in a two-layer quasigeostrophic model of the ocean around South Africa. They fed the model with gridded altimeter observations every 10 days but soon found that the relative weights varied too wildly. Only a few members had relatively large weights, while the rest had such low weights that it made no contribution in the posterior density to the first two moments. It was estimated that at least one million (!) ensemble members were needed to obtain statistically significant results in their application. A possible solution has been in fact known for quite a long time (Rubin 1988; Gordon et al. 1993): resample the posterior density after some time to create

a new ensemble in which all ensemble members have equal weight again. Clearly, the success of such a method depends on the density of the observations, but it is a promising candidate for the real solution. This resampling can be done in a variety of ways, and the results from importance resampling are discussed here. For an overview, the reader is referred to Doucet et al. (2001).

The new method versus importance resampling

The basic idea of importance resampling is extremely simple (see Rubin 1988). First, calculate the weight of each ensemble member. This collection of weights forms a new density. In the purest form of importance resampling, a sequence of numbers is sampled randomly from that density, in total, the amount of ensemble members N . Clearly, more numbers will be drawn at high weights than at low weights. The amount of numbers drawn at a certain weight is equal to the number of identical copies that are made of that ensemble member. So, if, for example, weight w_i is chosen four times, four identical copies of ensemble member ψ_i are made. On the other hand, if a weight is so low that it is not chosen at all, no copies of that ensemble member remain. In this way, a new ensemble, in which all ensemble members have equal weight again, is created. It is resampled from the density defined by the weights, and therefore by the relative closeness of the observations to the members. The procedure is depicted in Fig. 1.

In certain variants of importance resampling, the copies of an ensemble member are not identical, but some “jitter” is applied to obtain a little more spread in the ensemble. For instance, Anderson and Anderson (1999) use the kernel method (Silverman 1986), in which a Gaussian is created around each chosen ensemble member, and, instead of identical copies, new ensemble members are drawn from this Gaussian. The covariance structure of this Gaussian is, of course, problematic. In the standard kernel method, it is taken as a factor times the covariance of the whole ensemble, but the accuracy of this procedure is questionable. There is no a priori reason why the local structure of the probability density should resemble the global structure. Nevertheless, interesting results are obtained with several variants of the Lorenz63 model. In the application described here, no jitter is applied because the errors in the model dynamics are so large that identical copies will spread relatively fast. These dynamical model errors are added as random forcings in the model equations, so the model integrations are not deterministic.

The procedure that I followed in this paper closely resembles importance resampling, with the following modification. Instead of choosing randomly from the distribution determined by the weights, members with large weights are chosen directly from the distribution in the following way: First, this density is multiplied by the total amount of ensemble members. For each so-obtained weight that is larger than 1, the integer part of

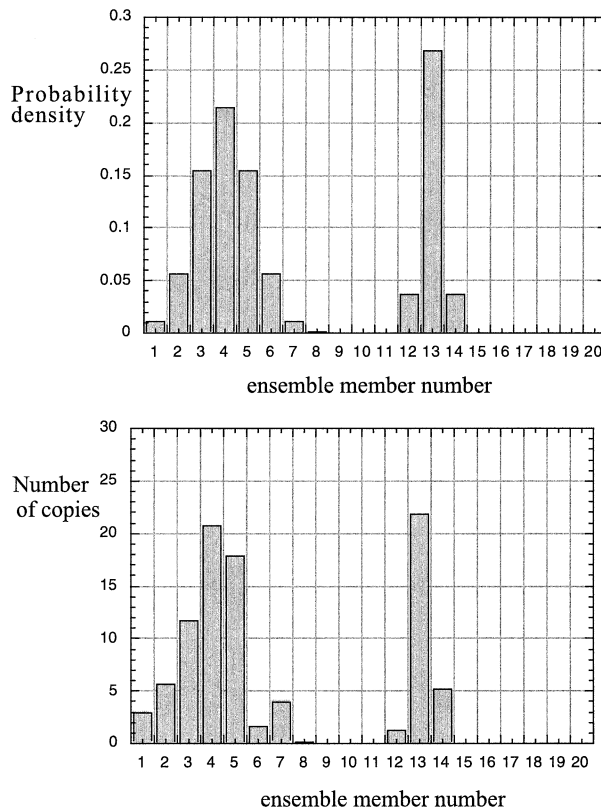


FIG. 1. The standard SIRF: (top) the posterior density (the weight of the individual ensemble members) before resampling and (bottom) the resampled ensemble, assuming an ensemble size of 100. The integer value denotes the number of copies of that ensemble member.

that weight determines the number of identical copies of the corresponding ensemble member. So, if the original weight of member i was $w_i = 0.115$, with an ensemble size of 100, the new weight $\tilde{w}_i = 100 \times 0.115 = 11.5$. This results in 11 copies of the ensemble member, while $\tilde{w}_i = 0.5$ remains. This procedure is followed for all weights. Finally, all remaining parts \tilde{w}_i form a new density from which the rest of the ensemble is drawn, according to the rules of stochastic importance resampling described above. The sampling method is demonstrated in Fig. 2.

The reason for the deviation from the basic importance resampling is as follows: Statistically, the basic rule is preferable if the ensemble is large enough. However, computational arguments lead to as small ensemble sizes as possible. In that case, we are more strongly interested in the members with high weights, and the above method minimizes the possibility that too many members are drawn with low weights. In this way, I try to maximize the amount of information present in the weighted prior ensemble on physical balances in the new ensemble. So, interestingly, while the posterior density has been sampled rather badly because of the low ensemble size, I keep to this sample as close as possible to avoid ensemble members that are useless from a sta-

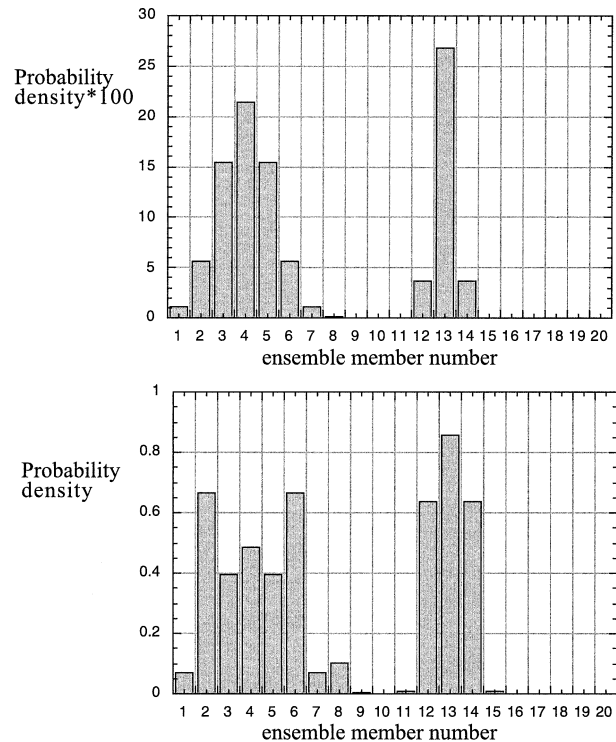


FIG. 2. The SIRF variant used herein: (top) the posterior density before resampling, multiplied by the ensemble size (100), and (bottom) the modified posterior after subtraction of the integer values of the top panel. The rest of the ensemble is drawn from this density.

tistical point of view. A question that comes to mind is why the second stage of the algorithm, in which the sampling is stochastic again, is not also done deterministically. The reason is that the stochastic sampling is so much simpler to implement. Deterministic sampling could be used again along the lines described above, but then a new density would arise. Also, this density could be sampled deterministically, but, again, a new density would arise, and so on, until the sum of the weights is smaller than 1. To avoid this process of deterministic sampling again and again, the stochastic method is used in the second stage.

From now on, this method is referred to as sequential importance resampling (SIR) to avoid further confusion in the data assimilation literature. However, the reader should take the differences with the standard method of that name into account.

A smoother version of the method can be obtained by using the weights at all times in the smoother time interval. The potential problem that only a few ensemble members are retained after a few filter steps, so that the smoother is based on a much smaller ensemble, needs further investigation and is postponed to a future paper.

A final comment on the particle nature of the filter and the resampling is that, because of these two approximations, the filter is not exactly variance minimizing: only the mean of the continuous probability

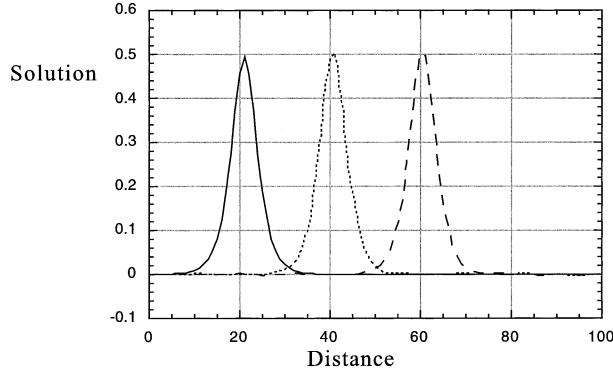


FIG. 3. True evolution of the soliton at $t = 0$ (solid line), 10 (dotted line), and 20 (dashed line). The measurements are taken around the max value (see text for details). Note the constant shape of the soliton. The horizontal axis denotes the grid points, with $\Delta x = 0.5$.

density has that privilege. However, because the continuous density cannot be calculated or stored in practice, the particle filter is as close as we can get. This closes the description of the data assimilation method used in this paper. In the next section, the filter version is applied to a simple tractable problem to compare its performance with standard methods like the ensemble Kalman filter.

3. A simple test case: The KdV equation

In this section, the new method is applied to the Korteweg–DeVries equation to study its behavior in nonlinear systems. For comparison, the ensemble Kalman filter is applied to the same problem. In this way, the difference between the conventional Kalman update and a variance-minimizing solution can be investigated.

The KdV equation describes the nonlinear evolution of a field u subject to advection and dispersion:

$$u_t + 6uu_x + u_{xxx} = 0. \quad (9)$$

[In fact, several forms of the KdV equation exist (see, e.g., Drazin and Johnson 1989).] Shape-conserving solutions called solitons are allowed by a balance between the steepening of the waveform due to nonlinear advection and the dispersion from the third spatial derivative. We start by a field of the form:

$$u(x, 0) = \frac{0.5a}{\{\cosh[\sqrt{a}(x - x_0)]\}^2}, \quad (10)$$

in which x_0 is the position of the maximum of the waveform, and $0.5a$ is its amplitude (see Fig. 3). The KdV equation will move the waveform toward positive x values with a speed a , while conserving its shape. Important for the following is that the soliton is stable to small perturbations (see, e.g., Drazin and Johnson 1989).

Several experiments have been performed with the new filter and the ensemble Kalman filter. Presented in detail here is one experiment that highlights the differences between the two methods. We first form a true

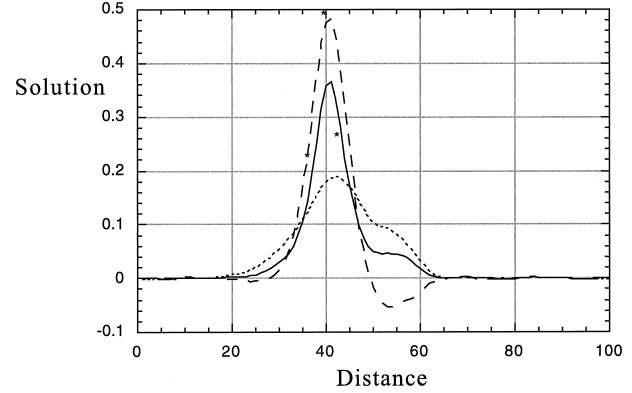


FIG. 4. Mean of the prior ensemble at $t = 10$ (dotted line), the mean of EnKF analysis (dashed line), and the mean of the SIRF analysis (solid line). The observations are indicated by stars. Although the EnKF solution is close to the observations, it is not optimal (variance minimizing). The SIRF solution uses the whole prior density and not only the first two moments. It is variance minimizing, so the model dynamics are taken into account consistently. Note the negative tail in the EnKF solution, leading to a breakup of the soliton later on.

solution by integrating this form with $a = 1$ over a domain of length 50, with periodic boundary conditions, $x_0 = 20$ and $\Delta x = 0.5$. The time-stepping scheme is leapfrog, with an Euler step every 67 time steps to suppress the numerical mode. The time step is $\Delta t = 0.1$.

This solution is measured six times, on $t = 10$, at $x = 37$, $x = 40$, and $x = 43$, and on $t = 20$, at $x = 57$, $x = 60$, and $x = 63$. To these pseudo-observations, random Gaussian noise with zero mean and standard deviation 0.05 was added. Note that the observations are taken around the peak values of the waveform.

We try to recover this solution with the sequential importance resampling filter (SIRF) and with the ensemble Kalman filter (see Burgers et al. 1998) for comparison. To this end, an ensemble was created with the amplitudes of the ensemble members drawn from a Gaussian with zero mean and standard deviation 0.5. To mimic errors in model dynamics, random numbers are drawn from a Gaussian distribution with zero mean and standard deviation 0.001. These numbers are added to the solution at each time step. The ensemble size was 250, but the results converged at $N = 150$. This ensemble was integrated forward in time, and data were added at $t = 10$ and $t = 20$, as explained above. Figure 4 gives the mean of the forecasting ensemble at $t = 10$. The decrease of the amplitude can be attributed to the spread in amplitudes a , leading to an ensemble of solitonlike waves with different propagation speed.

In the same figure, the mean of the ensemble after analysis at $t = 10$ is given for the ensemble Kalman filter and for the sequential importance resampling filter. The measurements are indicated by stars. The first thing that strikes the eye is that the EnKF solution comes much closer to the observations than the SIRF. However, the EnKF solution is not variance minimizing and is, in fact, too close to the observations. The EnKF assumes

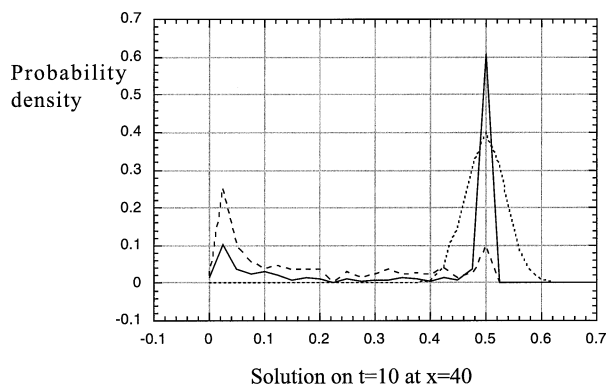


FIG. 5. The prior (dashed line), posterior (solid line), and observational (dotted line) probability densities on $t = 10$, $x = 40$ (the top of the soliton). The prior is not Gaussian, and neither is the posterior density, because of the nonnegative solution. The large variance in the prior is the reason that the EnKF solution becomes too close to the observations.

that the prior probability density of the model is Gaussian, but that is not the case. In Fig. 5, the prior, posterior, and observational density are given at the peak of the true soliton, on $t = 10$, at $x = 40$. The densities are created using the frequency interpretation on prespecified intervals. Varying the intervals within reasonable bands showed that the visible features are robust. Clearly, the prior is non-Gaussian because the solitons are always nonnegative. Several ensemble members moved too slow or too fast to have a significant value for the solution at $x = 40$. The variance of the prior is 0.3, which is more than a factor 10 larger than that of the observations. So, indeed, the EnKF posterior solution, being a weighted mean between the prior ensemble mean and the observations, has to be very close to the observation. If the prior is not Gaussian distributed, as is the case here, the posterior is not only determined by the mean and the variance of the prior and the observations but by the whole density. The variance-minimizing solution is the mean of the posterior density, which gives much more credit to the prior model estimate. The SIRF does give exactly this estimate. (Note that it has converged for 250 members.) So, the fact that the model is drawn close to the observations in the EnKF does not automatically mean that the EnKF analysis is good!

Variance estimates for a truly variance-minimizing solution also show unfamiliar behavior. Figure 6 shows the prior and posterior variance estimates for the SIRF at $t = 10$. Interestingly, the posterior variance is higher than the prior variance at the measurement point $x = 40$. So, in a conventional way of thinking, the uncertainty in the estimate at that point is increased because of the measurement. A more accurate inspection of the full prior and posterior densities shows that the uncertainty has decreased, but the second moment of the density does not show it. A way to quantify this uncertainty

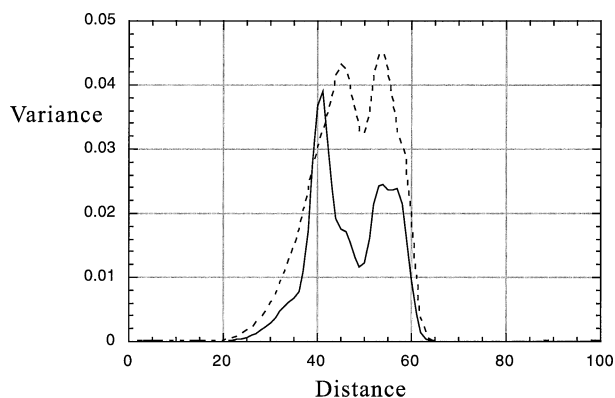


FIG. 6. Prior (dashed line) and posterior (solid line) variances on $t = 10$. The posterior variance is lower everywhere, but not at the measurement point at $x = 40$. Clearly, the variance is not a good measure of the accuracy in this case.

is by introducing the so-called entropy of the system as (see Shannon 1948)

$$H = -\sum_{k=1}^M p_k \log p_k, \quad (11)$$

in which M is the number of bins with a probability density unequal to zero. One can easily show that the entropy of a Gaussian-distributed variable is proportional to its variance. For probability densities other than the Gaussian, this is not necessarily true. For instance, for bimodal densities the variance loses its meaning. In that case, the entropy is a more sensible estimate, as Shannon (1998) showed. In our case, we find that the entropy of the prior density is $H = 1.1587$, and that of the posterior density is $H = 0.7286$. So, the entropy has decreased. We will come back to this in the next section.

We now turn to the analysis at $t = 20$. Figure 7 shows the mean states of the EnKF and SIRF ensembles after analysis. The SIRF is very close to the observations this time. This has to do with the correct update of the en-

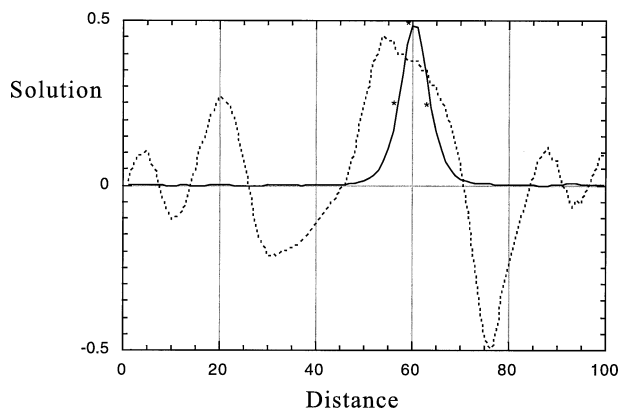


FIG. 7. Mean of the EnKF (dotted line) and SIRF (solid line) posterior solutions at $t = 20$. The stars denote the observations. The SIRF solution closely resembles the truth, but the EnKF solution shows a complete breakup of the soliton.

semble at $t = 10$, leading to a relatively large part of the ensemble rather close to the truth. The model error increases the spread in this ensemble a bit, so that the largest variance (about 0.045) is around $x = 60$ at $t = 20$. This variance is so large, and the ensemble is so close to the observations, that an almost perfect match is possible. The EnKF analysis looks rather strange. In fact, the program crashes some time after the update. Several reasons for this behavior can be found. First, Fig. 4 shows that a negative tail exists in the analysis of the EnKF. However, as is well known, negative values do give rise to fast wave motion toward the left in the KdV equation. The reason for this behavior is that the nonlinear advection and the dispersion enhance each other instead of balancing each other. This is indeed what happens with the updated ensemble members. They have negative parts, and also the perturbation due to the update is too large, so that the soliton falls apart rapidly. So, the mean is not evolving as a soliton rightward anymore, but it will break up. If one recalls that the figure shows the mean, one can imagine what the individual members must look like at analysis time, and how their subsequent evolution will be. The negative values arise because each new EnKF ensemble member is a linear combination of the old ones, without taking into consideration that no negative values exist, so without realizing that the prior is not a Gaussian. The resulting overfitting at the measurements leads to negative values farther on, because of the large gain. Another reason for the wild character of the update is that the ensemble is way off. The prior variance is extremely high, leading again to a large gain all over the domain, resulting in the problems mentioned above.

As mentioned above, several experiments have been performed with the KdV equation, and, depending on the system parameters, the above differences between the EnKF and the SIRF are more or less pronounced. Since little new can be learned from them, they are not discussed.

A serious problem is the issue of ensemble collapse. When the weights are such that the relative weight of one member is much larger than that of all the others, only that member is present in the determination of the moments of the posterior density. So, a mean will be produced, equal to that member, but the variance is close to zero. The filter thinks it is doing perfect, but that is not necessarily the case. The resampling will produce N (the size of the whole ensemble) identical copies of that member, consistent with the weighted ensemble, but again (of course) having the same problem. (Generally, the filter will diverge from the true evolution, so ensemble collapse is also termed filter divergence. Since divergence is one of the possible consequences of the collapse, we use the term ensemble collapse throughout this paper.) The experiments showed that the SIRF is more sensitive to ensemble collapse than the EnKF (with perturbed observations). On the other hand, when the EnKF collapses, it will be extremely difficult to get it

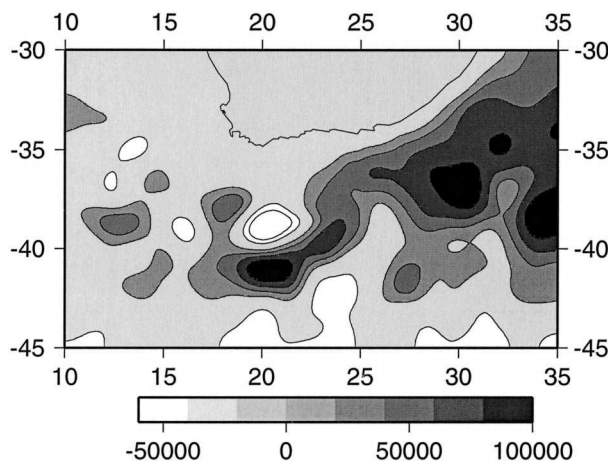


FIG. 8. Sea surface height field ($\text{m}^2 \text{s}^{-1}$) of the ocean area around South Africa ($30^\circ\text{--}45^\circ\text{S}$, $10^\circ\text{--}35^\circ\text{E}$), obtained from a snapshot of the ocean model. Note the Agulhas Current along the east coast of Africa, the shedding of Agulhas rings at the southern tip of the continent, and the strong meanders in the Agulhas return current flowing back east from the shedding area.

spreading again because the variance in the ensemble remains very low. The observations are unable to pull the ensemble to the correct state because their variance will generally be much larger. For the SIRF, the situation is different. The variance in the ensemble is not directly related to the weighting with the observations via the Kalman gain; instead, each member is weighted individually. The best member in the SIRF will be duplicated, and the new ensemble will spread again during the integration. The probability that more members are close to the next observation set increases strongly this way. As a result, the SIRF can be expected to be pulled more easily on the correct track than the EnKF. Crucial in the collapse problem is the choice of the probability density for the observations. The next section, which deals with a real application, elaborates further on this.

It should be mentioned that modifications to the standard EnKF can be used to overcome the problems presented here. For instance, one could put all negative values after analysis to zero, with or without compensation for the “mass loss.” Also, one could perform a local analysis to avoid the problems. This is almost always used in large-scale problems to prevent spurious correlations from destroying the solution. So, several relatively simple fixes are possible to make the EnKF work in practice. The fundamental point is, however, that the EnKF is not variance minimizing. The new method proposed here does not have this kind of problem by construction.

4. A real application

In this section, the new method is applied to a large-scale problem to study its behavior in such a setting. We study the ocean area around South Africa (Fig. 8).

The Agulhas Current runs along the east coast of South Africa southward and retroflects just after leaving the continent at its most southern tip, back into the Indian Ocean. At the retroflexion point, large Agulhas rings are shed, moving into the South Atlantic [see de Ruijter et al. (1999a) for a review]. The area is an ideal test bed for data assimilation methods because of the highly nonlinear dynamics and the availability of high quality satellite altimeter measurements of the height of the sea surface. Since this height is directly related to the pressure, it is an important dynamical constraint on the flow.

The area is modeled by a five-layer quasigeostrophic ocean model with a horizontal resolution of 10 km. The first baroclinic Rossby deformation radius is about 35 km in this area, so the model is so-called eddy permitting. This resulted in 251×140 grid points in the horizontal. The layer depths are 300, 300, 400, 100, and 3000 m, respectively, with densities of 1026, 1026.9, 1027.25, 1027.65, and $1027.75 \text{ kg m}^{-3}$. The time stepping was done with leapfrog, with an Euler step every 67th step to suppress the computational mode. A time step of 1 h was chosen, close to the Courant–Friedrichs–Lewy (CFL) limit, for optimal accuracy. Small-scale noise was reduced by a Shapiro filter of order 8. The quasigeostrophic dynamics do not allow an accurate description of coastal processes, so the continental shelf was assumed to be part of the land. From satellite infrared observations, the Agulhas flows along the continental shelf break, so that this assumption is probably not too severe. Boundary conditions are such that features are leaving the domain with twice the speed of the fastest wave mode. Only the inflow of the Agulhas Current at the eastern side of South Africa was prescribed. Because this inflow is supercritical for all baroclinic (Kelvin) waves, this last condition is well posed for those waves. Problems with barotropic waves did not arise. (Note that the values at the boundaries are part of the data assimilation problem.) The model is able to produce realistic ring-shedding events and general mesoscale motion in the area, as compared to satellite observations and in situ ship measurements.

a. Statistics

The initial streamfunction error (uncertainty) was taken space independent, with values of 4000, 3000, 2000, 1000, and $1000 \text{ m}^2 \text{ s}^{-1}$ for the layer models. Every day, a random error of 0.05 times these values was added to describe the model error. The spatial correlation of the errors was Gaussian with a decorrelation length of twice the Rossby radius of deformation. The value of the initial errors was rather low because the model was initialized from an interpolated altimeter sea surface height field. The time-mean field is always problematic because it is not well known. We used a field similar to that used by van Leeuwen (2001), but now interpolated over five layers. In this application, we added the same time-mean field to all observations, leading to consistency between

initial conditions and observations, but the real world might be inconsistent with this mean field. This will be visible in the data assimilation system by a bias in model dynamics (apart from the bias due to quasigeostrophy). However, as explained by van Leeuwen (2001), a bias does not prevent us from using the data assimilation equations, because they are still valid.

The state space, consisting of the five streamfunction fields for the five layers, has a dimension of about 2×10^5 . The ensemble size was 495. Extending this to 1024 did not alter the results significantly; that is, the change in the ensemble means was much less than the ensemble standard deviations, and the variance plots were identical visually.

b. Observations

The observations were satellite altimeter height data from the TOPEX/Poseidon (T/P) and *European Remote Sensing (ERS)-2* satellites. These two satellites cover the model area with tracks that are about 150 (T/P) and about 70 (ERS-2) km apart. While T/P has a repeat orbit of 10 days, ERS-2 has a repeat orbit of 35 days. The along-track resolution is 7 km.

A problem with satellite altimeter data is that the time-mean signal contains information about both the ocean circulation and the geoid. Since the geoid is not well known at the length scales of interest, only the time-varying part of the altimeter signal can be used. The time-mean field has to come from other sources. Here, we used a field derived from in situ measurements, as is also used in van Leeuwen (1999, 2001). Unfortunately, the accuracy of this field is not well known. [One could try to estimate the time-mean oceanic field by using it as the unknown in a data assimilation experiment, because the time-mean and the time-varying part of the signal are dynamically coupled. This is done by van Leeuwen (1999), but, although the results are encouraging, the quality of the field is still questionable. An independent estimate would definitely be preferable, and we hope that the Global Ocean Circulation Experiment (GOCE), which will determine the shape of the geoid, will help us out on this problem.]

For the initial field, an interpolated image was produced from the observations over a time period of 35 days, representative of the oceanic situation of 1 January 2000. The observations that are used in the data assimilation experiment are collected over 1 day, and the resulting batches are offered to the model. So, each batch has only a partial coverage of the domain, a few tracks, which differs from day to day. Every fifth observation was used in the data assimilation experiment, while every fifth observation with an offset of 2 was used as independent data to check the results from the assimilation experiment.

The observational error was specified as $5000 \text{ m}^2 \text{ s}^{-1}$, which corresponds to about 4 cm in sea surface height. This value is probably a bit too high for T/P data (2

cm) but a little too low for the *ERS-2* data (5 cm). The universal value was chosen here for simplicity, and because both data sources suffer from a not-well-defined time-mean sea surface topography. Recall that the purpose here is to demonstrate the abilities of the data assimilation system not the best reproduction of the oceanic state.

The shape of the probability density of the altimeter observations is a difficult matter. Because of the weighting procedure, the SIRF is very sensitive to the tails of that density. In general, we know little of those tails. It has been suggested, however, that the tails of a Gaussian are too small: the square in the exponent cuts off large deviations from the observations very drastically. So, outliers, meaning bad measurements here, may have a tremendous effect on the behavior of the filter. Indeed, the first experiments with Gaussian densities for the observations led to ensemble collapse directly at the first analysis time. Even increasing the observational errors by a factor of 10 did not help much.

This formed the motivation to look for densities with larger tails. The Lorentz density is used in this paper, but better alternatives can probably be found quite easily. The density is given by

$$f_d(\mathbf{d}|\psi) = \frac{1}{1 + \frac{[d - H(\psi)]^2}{\sigma^2}} \quad (12)$$

for uncorrelated observational errors. An advantage of this density is that it has a shape very similar to a Gaussian near the peak (symmetric and quadratic), but it is much broader away from the peak. The observational error is taken equal to σ , half the full width at half maximum, so equal to a Gaussian in this respect. The similarity with the Gaussian close to the peak is an important reason to use this density, although other choices might be just as good. A disadvantage is that the density has infinity variance, but that is only a theoretical problem not a practical one in this case.

c. Implementation

The SIRF was implemented on an Origin 2000 parallel computer using 15 processors. The distribution over the processors was done using the OpenMP multi processing programming system, in which each processor integrated 33 ensemble members. The analysis was done serially, but because no inversions had to be performed the code remained extremely parallel. The speedup was close to 100%, while the f90 code is extremely simple. The straightforward observational dataset required no special treatment. However, for a more involved assimilation parallel input/output (IO) seems to be in order. When more processors are present a message passing interface (MPI) version of the code is favorable because of communication problems in OpenMP with a large number of threads, but also be-

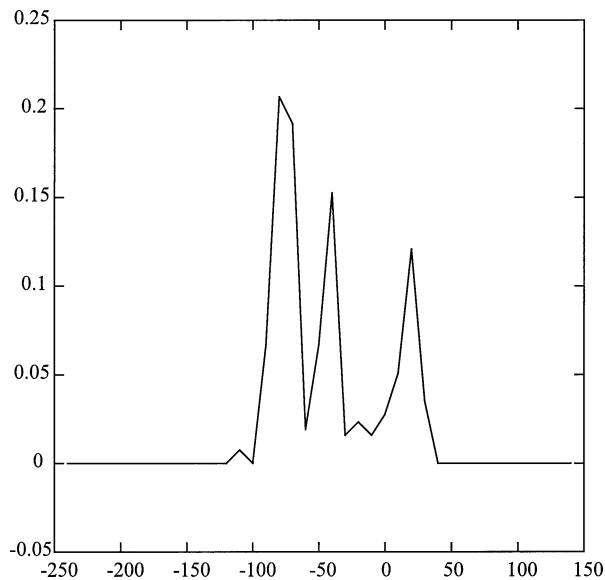


FIG. 9. Probability density of upper-layer streamfunction at a point in the retroreflection area. The density is non-Gaussian, so a Kalman filter cannot be used.

cause the MPI f90 code remains simple, with little extra coding compared to a serial version of the program.

d. Results

The carefully chosen error covariances lead to a sensible solution. First, the statistics of the results are discussed, including a comparison with independent observations; then, a short physical discussion of the obtained results is given.

1) STATISTICS

First, we have to make sure that a truly variance-minimizing method is needed, instead of a Kalman filter-like approach. Figure 9 shows the prior density of the upper-layer streamfunction at a point in the retroreflection area. The densities are created using the frequency interpretation on prespecified intervals. Varying the intervals within reasonable bands showed that the three visible peaks are robust. Clearly, this does not represent a Gaussian, as several statistical tests prove (not shown). So, a filter that goes beyond the Gaussian assumption, and so beyond Kalman filters, is needed for this problem.

Figure 10 shows the evolution of the total rms error of the ensemble for each model layer when the SIRF is applied. The rms is defined here as the rms deviation from the ensemble mean, since the truth is not known. The original large uncertainty is reduced strongly at the first measurement time, to stay close to this level for the rest of the assimilation period. Since the method weights each ensemble member as a whole, the error evolution in the lower layers is just a constant fraction

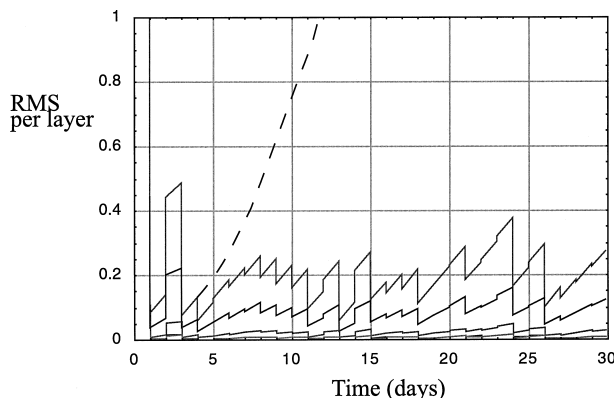


FIG. 10. Layer-integrated rms errors vs time. The values have been normalized by the error of the upper layer at day 1. All layers show a similar evolution because ensemble members are weighted as a whole. The observations are able to constrain the model evolution. Note that the variance occasionally goes up at analysis times (see text). The dashed line shows the upper-layer rms error when no observations are assimilated after day 3.

of that in the upper layer (represented by the top curve in Fig. 10). The dashed line that starts at day 4 shows the unconstrained evolution of the rms error of the upper layer when no observations are assimilated after day 3. It shows that the error growth is indeed more or less exponential, as one would expect from such a nonlinear model.

A feature that first attracts the eye is that the error occasionally goes up at analysis time. For Kalman filter-like methods, in which the assumption is made that model and observational prior densities are Gaussian, this would point to serious problems. The reason is that, when combining two Gaussian-distributed variables, the variance of the posterior density always has to be lower than either of the prior variances, as can easily be shown. For variables that are not Gaussian distributed, this does not have to be the case.

This fact, already demonstrated in the previous section, is illustrated here with a simple example. Consider two variables on the domain $[0, 1]$, with one having a density that is $4/3$ in $[0, 0.5]$ and $2/3$ in the rest of the interval, and the other having a density that is $2/3$ in the first half and $4/3$ in the second half of the interval. Obviously, the posterior density is 1 along the complete interval, and its variance is larger than that of either of the prior densities. In fact, this example is a bit depressing in the sense that also the entropy grows. So, the information in the observation is counteracting that in the model and shows that we should be less certain than the model predicts.

In the present case with the quasigeostrophic model, the entropy as measure of our uncertainty is calculated at the point corresponding to the probability density in Fig. 9. The entropy of the prior probability density is $H = 0.875$, while that of the posterior is $H = 0.731$. So, indeed, the entropy has reduced here too. As an informal statement, one might say that when the model

probability density is multimodal, and the observations favor only a few modes, the entropy, and therefore the uncertainty, will decrease, but the variance can increase because of the shifting mean value. No doubt more can be said on this, but that is beyond the scope of the present paper.

Another interesting feature is the fast initial drop of the error to a relatively low value, and the nearly constant value afterward. Kalman filter-like methods tend to show a more gradual decrease of the error. The reason must be that only a few ensemble members are close to the observations initially, so that only those members get a nonzero relative weight. The spread in those few members is relatively low. The resampling step then draws the complete ensemble to that part of state space.

Figure 11 shows the upper-layer rms error fields determined from the ensemble for $t = 0, 10, 20, 30$ days. The initial error is taken as $4000 \text{ m}^2 \text{ s}^{-1}$, corresponding to about 3 cm. The errors drop to values of about $1000 \text{ m}^2 \text{ s}^{-1}$, with occasional values of over $3000 \text{ m}^2 \text{ s}^{-1}$ after 10 days. On day 30, the error field has become more smooth, settling around $1500 \text{ m}^2 \text{ s}^{-1}$, or close to 1 cm. This seems unrealistically low, given the observational error of close to 4 cm. To investigate this further, the SIRF results were compared to independent observations.

In Fig. 12, the variance between the unused altimeter observations and the mean of the ensemble are compared to the variance of the ensemble at those observation points. The comparison shows that the ensemble spread is indeed what it should be. We thus can conclude that the SIRF is doing quite a good job.

2) PHYSICAL RESULTS

In Fig. 13, the mean of the analyzed ensemble is given for $t = 0, 10, 20, 30$ days. Because the errors are rather low, less than 10% in most of the area, the features that are visible have to be dynamically significant. A first thing that strikes the eye is the dipolar structure in the retroflexion area. The anticyclonic anomalies seem to wrap around the cyclonic feature. A ring is pinched off at day 20 but is recaptured at day 30. Meanwhile, the dipolar structure is moving westward with a speed of about 10 km day^{-1} , or close to 12 cm s^{-1} . Farther west, at $38.5^\circ\text{S}, 12.5^\circ\text{E}$, an Agulhas ring is present at day 0. At day 30, its center has moved farther west, with a speed of close to 3 cm s^{-1} . This is within the observational range from a large collection of these rings (see, e.g., Schouten et al. 2000). It might be interacting with a similar ring farther southeast, but that is not conclusive from the figure. Also interesting is the cyclonic feature between the retroflexion and the Agulhas ring. It seems to vanish at day 20 and to disappear completely at day 30. This cyclone might be the result of a spinning up of the water mass due to the presence of anticyclonic features at its eastern, southern, and western sides. Its fast disappearance might point to the fact that the cy-

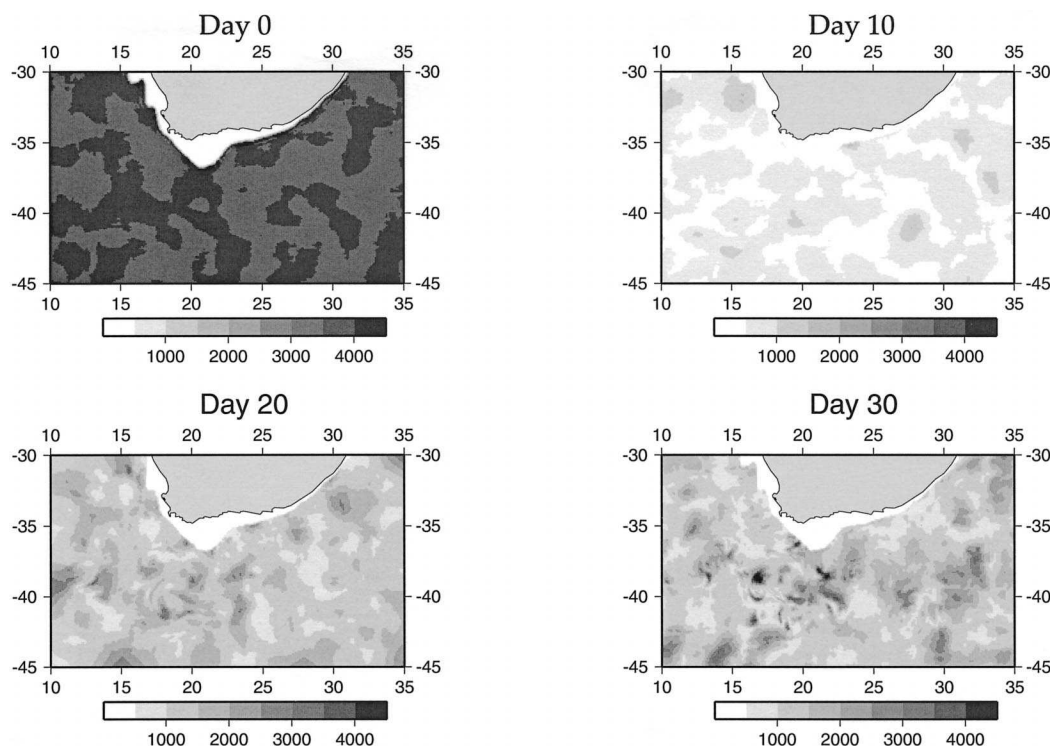


FIG. 11. Root-mean-square error fields ($\text{m}^2 \text{s}^{-1}$) for the upper layer for days 0–30. The errors are reduced more or less uniformly because the ensemble members are weighted as a whole.

clone is not a coherent feature. [Cyclonic rings of this kind have been observed before in situ data (S. Garzoli 1999, personal communication).]

On the other hand, the cyclonic part of the dipole does seem to be a dynamical entity. Indeed, while the anticyclonic rings are more likely to be formed at a retroflection of this kind, the formation of cyclonic lee eddies at the point where the Agulhas Current separates from the continental shelf has been documented recently (J. Lutjeharms 2001, personal communication). The cy-

clonic eddy might be amplified by cyclonic meanders in the Agulhas Current, so-called Natal Pulses, that form close to Durban, South Africa, at 30°S , and move downstream toward the retroflection area (see Lutjeharms and Roberts 1988; de Ruijter et al. 1999b; van Leeuwen et al. 2000). More interesting flow features can be observed, such as the pinching off of a cyclonic eddy in the Agulhas return flow, near 37°S , 32°E .

From this discussion, it becomes clear that the area is governed by highly nonlinear dynamics, so that a data assimilation method that goes beyond Gaussian statistics might be needed. I elaborate on this in the concluding section.

5. Summary and discussion

A new data assimilation method, a variant of sequential importance resampling, has been presented. It is truly variance minimizing for nonlinear dynamics, unlike schemes based on the (ensemble) Kalman filter. It is argued that because of its very nature, data assimilation is not an inverse problem, or a minimizing problem, although it is usually cast in that form. The method is based on importance resampling, a well-known method in control theory. New is the application to real-size problems in this paper.

The method was tested on the KdV equation, and its behavior was compared to that of the ensemble Kalman

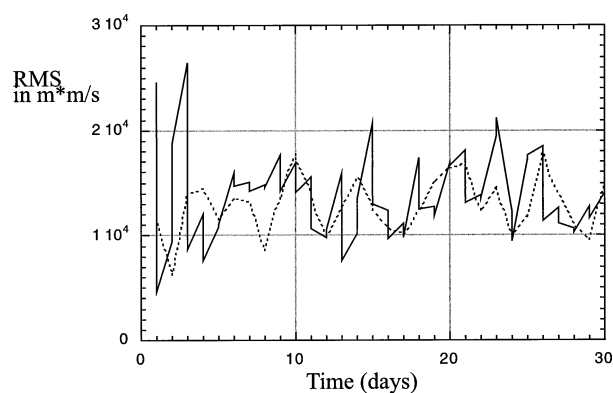


FIG. 12. Comparison of rms error ($\text{m}^2 \text{s}^{-1}$) between ensemble mean and independent observations (dotted line) and the std dev in the ensemble (solid line). The excellent agreement shows that the SIRF is working correctly.

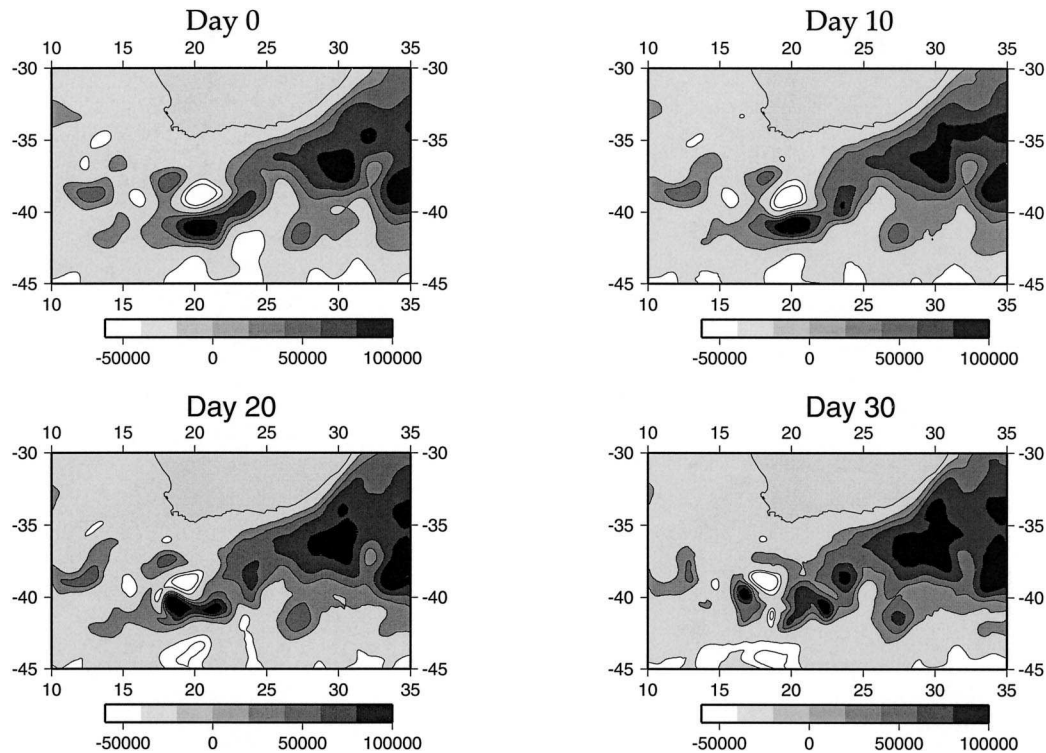


FIG. 13. Mean of the analyzed SIRC streamfunction fields ($\text{m}^2 \text{s}^{-1}$) of the upper layer. Note the dipolar structure south of the continent that is dominating the ring-shedding process.

filter. It was shown that the new method leads to new filter behavior, uncommon to Kalman filter methods. For instance, Kalman filters might be drawn too close to observations because of the implicit Gaussian assumption of their prior density. It should be mentioned that the actual prior and posterior densities in ensemble Kalman filters can be highly non-Gaussian; the Gaussian assumption is only made at analysis time. Furthermore, it is possible that the variance of the posterior density is larger than that of the prior density at some locations, as the new filter showed. This is impossible when the Gaussian assumption is made. Inspection of the whole density showed that it did contract toward the observations, so the overall uncertainty is reduced. Finally, it was shown that the new filter avoids unbalanced states. The new method is sensitive to ensemble collapse when only a few members are relatively close to the observations, even more so than the EnKF. However, several experiments with the KdV equation showed that the filter is able to recover if the model noise is large enough, probably more easily than the EnKF. Although several experiments have been performed, a more precise statement cannot be given at this moment.

The new variance-minimizing filter was tested with a large-scale problem. Because of the high dimension of the state space in real-size applications, care has to be taken when formulating the probability density of the observations. It was argued that because of outliers, a Gaussian density is too narrow; that is, its tails are

too low. One measurement can degrade the solution if this is not taken into account. A Lorentz profile was used in this paper, leading to a more robust filter that seemed to work quite well. A comparison with independent observations showed that the spread in the ensemble closely resembled the true rms error of the solution.

An interesting feature of the method is that the variance of the posterior density can be larger than that of the prior density. This can be understood in terms of the non-Gaussianity of the method. It can be shown that a measure of uncertainty from information theory, the entropy, is decreasing in the examples presented in this paper.

To sum up, the new method is truly variance minimizing, no matrix inversions are needed, the observations can be distributed non-Gaussian, and the measurement functionals can be nonlinear without any problem. Furthermore, it preserves prior model constraints such as positive definiteness, unlike Kalman filter-like methods that mix states at analysis time, and provides error estimates, unlike four-dimensional variational data assimilation (4DVAR)-like methods. Finally, it is easy to implement and parallel, by its very nature.

Notwithstanding these advantages, some challenging issues remain. The matter of ensemble collapse has to be investigated further. It might be related to the true dimension of the attractor, the quality of the observations, the initial error estimates and the first guess itself,

the errors in the model dynamics, etc. Only practical experience will probably help us here. Recently, several variants of the EnKF have been proposed in which a forgetting factor is introduced in the Kalman gain [e.g., the singular evolutive extended Kalman (SEIK) and singular evolutive interpolated Kalman (SEIK) filters; see Pham (2001)]. The procedure requires that a factor be placed in front of the error covariance of the observations in the Kalman gain, and that the factor be smaller than 1. Another related method, in which a covariance inflation factor is used, is proposed by Anderson (2001). This factor multiplies the prior model density to prevent ensemble collapse. The result of both methods is that the error covariance of the observations is artificially reduced compared to that of the model, so that the model state is pulled more to the observations than it should be, according to the Kalman filter. (So, the filter partly forgets the prior model state.) This will prevent filter divergence from occurring. Indeed, much smaller ensembles can be used without ensemble collapse [Pham (2001) can follow a trajectory on the Lorenz63 attractor that is only partially measured with only three ensemble members], but it is unclear what is actually done. One can argue that the Kalman filter is not optimal for nonlinear problems, so that some freedom is present. On the other hand, it pulls the analyzed fields closer to the observations, but that is not what data assimilation is about in the first place. To put it to an extreme, the model should not be used as an interpolator between the observations; it contains important information in itself.

Another problem with the SIRF is that it performs only a reweighting of the ensemble members, so no “new blood” enters the ensemble at analysis times. The EnKF has the same problem: each updated member is just a linear combination of the old ones. The reason that the EnKF works so well for even a fairly large state space (over one million; the DIADEM group 2000, personal communication) is that the updates are done locally. Apart from decreasing the noise in the estimates, as Houtekamer and Mitchell (1998, 2001) argue, it also updates the large-scale structure of each ensemble member. A straightforward extension of the present method is to implement local updating. A problem is then that, at some point, the solution will jump from one member to the other, potentially leading to strongly unbalanced states. EnKF-like methods do not have this problem because of the smearing effect of the Kalman gain.

One can question why the method can work in a dynamical situation with many growing modes. For example, if the number of growing modes is larger than the number of ensemble members, the method cannot work. However, this is only true if the modes do not interact. We know from quasigeostrophic dynamics (see, e.g., van der Vaart et al. 2001) that growing modes have strong nonlinear interactions, producing new unstable modes and rectifying main flows. So, just counting the number of growing singular vectors might be mislead-

ing. The truly interesting variable is the dimension of the attractor of the model, but to calculate that is still beyond our power. Evidence that the quasigeostrophic model used here is indeed highly nonlinear is given by the rapid error growth in the unconstrained run and the highly non-Gaussian probability densities presented in Figs. 9 and 10. This discussion shows that even if methods do not work on simple nonlinear models, they might work in more realistic settings like a multilayer quasigeostrophic model. Obviously, much more research is needed to push these statements beyond mere conjecture.

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: On the analysis scheme of the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- de Ruijter, W. P. M., A. Biastoch, S. S. Drijfhout, J. R. E. Lutjeharms, R. P. Matano, T. Pichevin, P. J. van Leeuwen, and W. Weijer, 1999a: Indian–Atlantic interocean exchange: Dynamics, estimation and impact. *J. Geophys. Res.*, **104**, 20 885–20 910.
- , P. J. van Leeuwen, and J. R. E. Lutjeharms, 1999b: Generation and evolution of Natal pulses: Solitary meanders in the Agulhas Current. *J. Phys. Oceanogr.*, **29**, 3043–3055.
- Doucet, A., N. de Freitas, and N. J. Gordon, Eds., 2001: *Sequential Monte Carlo Methods in Practice*. Springer, 581 pp.
- Drazin, P. G., and R. S. Johnson, 1989: *Solitons: An Introduction*. Cambridge University Press, 226 pp.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- , and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867.
- Gordon, N. J., D. N. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc.*, **140F**, 107–113.
- Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718–1728.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- , and —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- Lutjeharms, J. R. E., and H. R. Roberts, 1988: The Natal pulse: An extreme transient on the Agulhas Current. *J. Geophys. Res.*, **93**, 631–645.
- Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models using the ensemble Kalman filter with a quasi-geostrophic model. *Tellus*, **51A**, 167–194.
- Pham, T. P., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Rubin, D. B., 1988: Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics 3*, J. M. Bernardo et al., Eds., Oxford University Press, 395–402.
- Schouten, M. W., W. P. M. de Ruijter, and P. J. van Leeuwen, 2000: Translation, decay and splitting of Agulhas rings in the south-eastern Atlantic Ocean. *J. Geophys. Res.*, **105**, 21 913–21 925.

- Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 175 pp.
- van der Vaart, P. C. F., and W. P. M. de Ruijter, 2001: Stability of western boundary currents with an application to pulse-like behavior of the Agulhas Current. *J. Phys. Oceanogr.*, **31**, 2625–2644.
- van Leeuwen, P. J., 1999: The time-mean circulation in the Agulhas region determined with the ensemble smoother. *J. Geophys. Res.*, **104**, 1393–1404.
- , 2001: An ensemble smoother with error estimates. *Mon. Wea. Rev.*, **129**, 709–728.
- , and G. Evensen, 1996: Data assimilation and inverse problems in a probabilistic formulation. *Mon. Wea. Rev.*, **124**, 2898–2913.
- , W. P. M. de Ruijter, and J. R. E. Lutjeharms, 2000: Natal pulses and the formation of Agulhas Rings. *J. Geophys. Res.*, **105**, 6425–6436.
- Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.*, **11**, 349–368.